In response to the Copyright Office's Notice of Inquiry and Request for Comments, Docket No. 2023-6 :

Artificial Intelligence and Copyright

This response is submitted on behalf of the AI and Metaverse Task Force of the Trust over IP Foundation (ToIP). The Trust over IP Foundation is an independent project hosted at the Linux Foundation, working with pan-industry support from leading organizations around the world. Our mission is to provide a robust, common standard and complete architecture for Internet-scale digital trust. For more information: https://trustoverip.org/

Responses to Questions

Question 1.

As described above, generative AI systems have the ability to produce material that would be copyrightable if it were created by a human author. What are your views on the potential benefits and risks of this technology? How is the use of this technology currently affecting or likely to affect creators, copyright owners, technology developers, researchers, and the public?

Response to Question 1.

Generative AI systems operate based on the instructions provided by a human user. These systems lack consciousness and shouldn't be considered as authors or creators. At the heart of any creation facilitated by such systems is a human user harnessing a tool for a specific purpose.

For instance:

A programmer might employ Code Co-Pilot to automate the generation of code segments.
A game designer could use Midjourney for graphic design tasks or ChatGPT for crafting character dialogues. In the future, Generative AI tools might even assist designers in producing animated video clips.
In each of these scenarios, a human (be it a programmer or a game designer) is leveraging GAI towards fulfilling a creative objective.

Some have described the experience of using GAI tools as akin to collaborating with an assistant. However, it's crucial to differentiate between a tool and a human assistant. While there might be certain parallels, the distinctions are significant. Moreover, even if one were to equate the AI's role to that of an assistant, the content generated with the help of the AI tool would be akin to "work for hire." Thus, the intellectual property rights (IPR) of the resulting output rightfully belong to the human creator.

Question 2.

Does the increasing use or distribution of AI-generated material raise any unique issues for your sector or industry as compared to other copyright stakeholders?

Response to Question 2.
Copyright protection is fundamental for developers in the software and digital media sectors. In our highly digitized tech industry, computer tools are integral at every stage of software and digital media creation. Incorporating AI into this workflow is not new; however, the recent influx of advanced Generative AI (GAI) tools represents an evolutionary stride in automation and creative innovation within the field. Altering the existing copyright framework could send significant ripples through the industry.

In contrast to art, where an original piece by a master may be more valued than its digital counterpart, software is assessed based on its functionality, not its production method. The software industry is inevitably driven to utilize every available tool, including GAI, to deliver the most efficient software at the lowest cost. Thus, the extent of copyright protection afforded to software should remain unaffected by the tools employed in its creation.

Similarly, these principles apply broadly to digital media, ranging from today's digital photography to the intricate 3D photorealistic avatars anticipated for metaverse applications and beyond. Digital media, much like software, is valued for its end product, not its production method. The inherent characteristic of digital media is that its worth isn't determined by its creation process but solely by its final output. In this realm, one copy is indistinguishable from another, highlighting the non-discriminative nature of digital replication.

A fitting comparison can be drawn between how contemporary copyright protection extends from software source code to its corresponding compiled binary. In the realm of GAI tools like ChatGPT, the prompts or inputs can be analogized to the "source code." Meanwhile, the resultant text or image can be likened to the "compiled" binary output. Just as copyright protection encompasses both the source code and its compiled form, it should similarly cover the input prompts, source datasets, and the ensuing articles or images produced by GAI.

Question 8.
Under what circumstances would the unauthorized use of copyrighted works to train AI models constitute fair use? Please discuss any case law you believe relevant to this question.

Response to Question 8.
When we purchase a published book, the insights and knowledge we gain from reading it aren't bound by copyright restrictions. After all, the very essence of establishing copyright protection is to incentivize authors to write and publish, ensuring that society as a whole reaps the benefits of shared knowledge.

Copyright primarily guards against unauthorized "copying." It differs from a patent, which safeguards novel inventions or ideas. Profiting significantly from the knowledge acquired from a book is entirely permissible and, indeed, the desired outcome. The core question then becomes whether utilizing a book for AI training constitutes "copying" or simply "reading" and "understanding." This question delves deeper than merely determining if such usage qualifies as "fair use."

A Generative AI training program processes digital files of copyrighted materials, assuming they are in digital formats, but subsequently discards the original files, retaining only the computational results—primarily as parameters within the model. Not a single word or bit of information from the original sources is directly replicated within the AI model. Given these circumstances, one must scrutinize whether the transformative process within the training program truly represents an act of "copying."

We contend that such training programs, as typically executed in contemporary deep learning algorithms, should not be classified as "copying" for several reasons:

• Transformation of Form: The outputs of these training programs differ markedly from the original material. It's inconceivable for anyone to identify fragments or excerpts from a book regardless of how the parameters in the model are presented. The model does not house a copy of any book. Instead, it assimilates statistical data from the literature to understand the interrelation of words.

• Inability to Replicate Original Material: One cannot utilize the trained model to recreate substantial portions of a book. While one might prompt, for instance, ChatGPT, to reproduce a brief quotation from a book present in its training data, such a recreation aligns with the principles of fair use.

Question 15.

In order to allow copyright owners to determine whether their works have been used, should developers of AI models be required to collect, retain, and disclose records regarding the materials used to train their models? Should creators of training datasets have a similar obligation?

Response to Question 15.

Developers of AI models have a clear imperative: they should meticulously document, store, and reveal the materials leveraged in training their models. The rationale for this extends beyond mere copyright concerns. The pivotal advantages of such a practice encompass:

• Reproducibility: Maintaining a thorough record of training materials ensures that the model's results can be consistently replicated, affirming the model's reliability.

• Accountability & Verifiability: By disclosing the sources of training data, developers take responsibility for the content that informs their models. It is a step towards ensuring that the AI system's decisions are derived from trusted and reputable sources.

• Responsible AI Development and Use: Knowing the origin of a model's data is paramount. It underpins ethical considerations and aids in addressing potential biases, ensuring that the AI operates in a manner consistent with societal values and norms.

The call for transparency in AI training data isn't merely a footnote or an afterthought. It is fundamental to the comprehensive understanding, judicious regulation, and safety of AI

systems. The imperative remains regardless of how copyright laws might interpret this responsibility.

Question 17.
Outside of copyright law, are there existing U.S. laws that could require developers of AI models or systems to retain or disclose records about the materials they used for training?

Response to Question 17.
Absolutely, as outlined earlier, from the perspective of responsible and safe AI practices, there's a compelling case for mandating full documentation and transparency of datasets employed in training. This holds true regardless of any potential copyright considerations associated with the dataset. Ensuring transparency not only upholds the principles of accountability and reproducibility but also establishes trust, which is paramount in the age of rapidly advancing AI technologies.

Question 18.
Under copyright law, are there circumstances when a human using a generative AI system should be considered the ‘‘author’’ of material produced by the system? If so, what factors are relevant to that determination? For example, is selecting what material an AI model is trained on and/or providing an iterative series of text commands or prompts sufficient to claim authorship of the resulting output?

Response to Question 18.
The person using AI to make something should always be seen as the creator, even when their role may seem small. If we say AI isn't human and can't have rights, then it also can't be called a creator. We shouldn't give AI human traits sometimes and not others. This would only make legal matters messy and confusing.

People argue that if software like Midjourney creates every part of a digital image, then a human isn't the author. But this is like saying a photographer isn't an author because the camera captures every pixel, either using computer tech or chemicals. No human touch is there. But as we use more digital tools, we should remember that one digital bit isn't different from another. The real creator is the person with a vision who uses these tools, even if their direct input seems minimal.

Giving a "prompt" to create a picture is a way for humans to show intent and give direction. It's a creative act. Crafting a prompt can take a lot of effort, creativity, and time. This process should be viewed as creative, and the result should be seen as a copyrighted work by that human. Asking if text commands or prompts are enough is missing the point. Just like a photographer's single click doesn't capture the entire creative journey leading up to that shot, we shouldn't downplay the creative process behind a prompt. The effort before the final action is crucial in both cases.

We stand with the Writer's Guild of America (WGA) strikers in their assertion[1]. They believe that writers are the true creators, even when they utilize GAI in their work. These writers should benefit from the efficiencies AI brings to their craft.

Question 20.
Is legal protection for AI-generated material desirable as a policy matter? Is legal protection for AI-generated material necessary to encourage development of generative AI technologies and systems? Does existing copyright protection for computer code that operates a generative AI system provide sufficient incentives?

Response to Question 20.
We assert that materials produced using generative AI tools shouldn't be labeled as "AI-generated." Instead, they are crafted by humans using these AI tools, making them human-generated content. The responsibility and ownership of this content should rest with its human creator.

If we attempt to exclude the portion crafted by a tool from copyright protection, we face the daunting challenge of defining what constitutes an AI tool. This approach is problematic.

All AI systems are, at their core, just algorithms. Their advanced capabilities only demonstrate that intelligence can be emulated computationally. There's no true distinction between AI-driven computations and other types of calculations. A uniform approach to copyright law is essential. It would be detrimental if we were to award copyright for one algorithm but deny it for another.

Question 28.
Should the law require AI generated material to be labeled or otherwise publicly identified as being generated by AI? If so, in what context should the requirement apply and how should it work?

Response to Question 28.
We advocate for labeling the method of production and processing of material without explicitly mentioning AI. After all, various tools, such as Photoshop, have been utilized to modify digital content for years. AI merely represents a category of algorithms. Labeling, devoid of AI referencing, is useful and entirely feasible.

Many organizations and companies are already pioneering methods and tools for convenient labeling. For example, the Trust over IP Foundation[2] is crafting a data authenticity standard through its ACDC Task Force, and its Trust Spanning Protocol Task Force is formulating a standard protocol to transmit such authenticity information among parties. Another entity, C2PA[3], is formulating standards for media files, such as photographs, to document their creation provenance, thereby supporting authenticity that consumers can utilize to evaluate the veracity of a photograph. The use of an AI tool during the creative process can be considered factual metadata, distributed alongside the work, enabling viewers to comprehend the truth and evaluate accordingly.

---

[1] https://variety.com/2023/biz/news/wga-new-contract-strike-ai-writers-room-staffs-residuals-1235736648/
[2] https://trustoverip.org/
[3] https://c2pa.org/

Question 29.
What tools exist or are in development to identify AI-generated material, including by standard-setting bodies? How accurate are these tools? What are their limitations?

Response to Question 29.
There are two primary tools being developed to determine the use of AI in digital content. First, there are watermarking tools, which are easy to use but might not be the most reliable. Second, there are cryptographic tools, which allow creators to make verifiable claims about their content. We believe the latter offers the most potential.

The Trust over IP Foundation (ToIP) has the following ongoing work that can help address this challenge:

- ToIP AI and Metaverse Task Force

This task force is working on understanding the issues of digital trust in the age of Generative AI and proposing standards that can offer Internet scale solutions. Identifying material's origin and provenance is one area of study, the other is to tell apart a human agent and an AI powered agent.

This task force produced this response to the Notice of Inquiry.

- ToIP Trust Spanning Protocol Task Force

The Trust Spanning Protocol is a draft standard being developed to provide the foundational communication protocol that digital entities can be certain about the other party's identity and authenticity, including the identity of being human or an AI agent.

- ToIP ACDC Task Force

The Authentic Chained Data Container (ACDC) Task Force is drafting a family of specifications that define the standard requirements for the semantics of Authentic Chained Data Containers. The semantics of ACDCs include both source provenance, governance and authorization provenance or delegation.

In addition, the work of the Coalition for Content Provenance and Authenticity (C2PA) also falls on the similar approach.

- The Coalition for Content Provenance and Authenticity (C2PA)

The Coalition for Content Provenance and Authenticity (C2PA) addresses the prevalence of misleading information online through the development of technical standards for certifying the source and history (or provenance) of media content. C2PA standard defines a standard way to

assert provenance meta-data so that consumers can easily access such information to tell the nature of digital materials.

End of Responses